A JOURNAL BY SISSA MEDIALAB

JCom

SPECIAL ISSUE
Emotions and Science Communication

ARTICLE

Empathic, humorous, and…trustworthy? A mixed-methods study on real-time evaluations of voice-based AI communicating science-related information

Evelyn Jonas and Monika Taddicken

Abstract

Two studies, using real-time response measurement and interviews, explore how German recipients assess the trustworthiness of a voice-based communicative AI conveying science-related information with empathic and humorous expressions. In both a laboratory and an online study, humor was associated with short-term declines in trustworthiness, reflecting cultural expectations and appreciation of objectivity and neutrality. In contrast, empathic expressions are rated more trustworthy, but evoke less conspicuous effects. Cluster analysis identified four distinct patterns of evaluation, with two groups largely unaffected by affective cues and two skeptical of humor, underscoring the importance of personalization and adaptation in designing trustworthy communicative AI for science communication.

Keywords

Public perception of science and technology; AI tools in science communication; Digital science communication

Received: 30th April 2025
Accepted: 23rd September 2025
Published: 10th November 2025

8

1 ▪ Introduction

In 2024, 54% of regular ChatGPT users in Germany reported asking the chatbot for science-related information [Greussing, Guenther et al., 2025]. Even before ChatGPT's launch, a notable share of German users consulted voice assistants for complex decisions [Greussing, Jonas & Taddicken, 2025]. These findings suggest that communicative AI (ComAI), whether as chatbot or voice assistant, is evolving from a simple channel for human communication to a conversational partner [Guzman & Lewis, 2019] and intermediary for science-related information. Despite valid concerns about an AI-driven 'infodemic' [e.g. Jungherr & Schroeder, 2023], ComAI holds promise for combating misinformation and lowering access barriers to complex information [Gong & Su, 2024; Schäfer, 2023]. Given its growing relevance, understanding how users evaluate ComAIs trustworthiness becomes essential [Jonas et al., 2025].

While most users appreciate the clarity and structure of ComAI-generated text [Skjuve et al., 2024], science communication goes beyond functionality. Science communication researchers advocate for warmer, more affective approaches, like humor [Riesch, 2014] and empathy [Bray et al., 2012] to engage non-expert audiences. Similarly, developers are increasingly embedding affective expressiveness into ComAI [Concannon & Tomalin, 2023], using humor [Lopatovska, 2019] and empathy to counterbalance coldness and the absence of human touch, and facilitate trustworthiness [Seitz, 2024]. However, these strategies may conflict with users' expectations of both the communicator and the subject matter, complicating trustworthiness evaluations.

Empathy and humor have been studied in human-machine communication (HMC) research in areas like customer service or healthcare [e.g. Liu & Sundar, 2018; Shin et al., 2023], but their role in science communication remains underexplored. Science communication research has focused chiefly on human communicators. This paper addresses this gap by investigating how empathic and humorous responses from ComAI influence perceived trustworthiness.

Two exploratory studies were conducted in Germany, using guided interviews ($n_{\text{lab}} = 15$), real-time response (RTR) measurements, and standardized questionnaires ($n_{\text{lab}} = 36$; $n_{\text{online}} = 503$). They explore how non-scientists assess a ComAI's trustworthiness in real-time (RQ1) and the reasons behind their evaluations (RQ2). A hierarchical cluster analysis on the RTR data identifies distinct evaluation patterns and the role of perceived human-likeness, empathy, and humor (RQ3). The findings provide empirical insights and practical guidance for designing emotionally attuned, yet responsible AI-based science communication.

2 ▪ Perceptions of ComAI and the communication of science-related information

2.1 ▪ *Evaluating trustworthiness*

Trust is defined as the willingness of a trustor to be vulnerable, expecting the trustee to act beneficially without direct oversight [Mayer et al., 1995]. In human-AI interactions, (Com)AI's computational abilities enable it to identify patterns and draw conclusions from vast datasets, surpassing human capabilities. This creates an epistemic disadvantage for users, especially when verifying (Com)AI's responses independently [Seger, 2022]. Trust, in this context,

stems from users' epistemic dependence on a better-informed (Com)AI and the risk of misinformation [Hendriks et al., 2015; Seitz et al., 2022]. While trust reflects the trustor's behavior, trustworthiness pertains to the trustee's inherent quality, evaluated across various dimensions [Mayer et al., 1995]. Given ComAIs hybrid perception as both machine and human-like [Etzrodt & Engesser, 2021], a framework for trust in ComAI in science-related contexts incorporates machine-trustworthiness dimensions — functionality, reliability, and helpfulness [Mcknight et al., 2011] — alongside human epistemic dimensions [Hendriks et al., 2015], including expertise, integrity, and benevolence [Jonas et al., 2025].

Human-like interfaces are expected to enhance trustworthiness perceptions [Glikson & Woolley, 2020] — an assumption rooted in the "Computers are Social Actors" paradigm [Nass et al., 1994] and MAIN (Modality, Agency, Interactivity, and Navigability) Model [Sundar, 2008]. The former posits that users treat computers socially as if they were human-like, while the latter expands on this by describing how specific technological cues activate heuristics that shape trustworthiness perceptions. Alongside anthropomorphic cues like voice or gesture, ComAI's affective communication can promote the social presence heuristic, potentially increasing users' liking and trust [Glikson & Woolley, 2020; Sundar, 2008]. Against this backdrop, agency cues, like expressions of empathy and humor, can enhance interpersonal interactions and relational satisfaction [Hampes, 2010], boosting ComAI's trustworthiness by reinforcing perceptions of expertise, integrity, or benevolence [Brummernhenrich et al., 2025; Shao et al., 2025; Xie et al., 2024]. However, overly realistic cues may evoke discomfort or eeriness — known as the "uncanny valley" effect, though this diminishes with increasing familiarity [Złotowski et al., 2015].

At the same time, awareness of communicating with a ComAI can activate the machine heuristic, stereotypically associating ComAI with precision, objectivity, and unemotionality, therefore making it perceive as trustworthy [Sundar & Kim, 2019]. Thus, expressions of empathy and humor may create a tension between these heuristics. Applying this ambivalence to science communication, where science is often perceived as cold or robotic [Rutjens & Heine, 2016], requires a deeper understanding of the role of affective[1] (e.g., empathic or humorous) ComAI as an information intermediary, that combines both perspectives.

2.2 ▪ Empathy in science and human-AI communication

Decades of multidisciplinary research have made empathy a complex construct with numerous definitions [Cuff et al., 2014]. However, empathy is broadly recognized as central to interpersonal relationships [Concannon et al., 2023] and communication [Zhang & Lu, 2024]. It is commonly delineated into three dimensions [Clark et al., 2019]: (1) Cognitive empathy, understanding others' thoughts and emotions; (2) affective empathy, sharing those thoughts and emotions congruently; and (3) behavioral empathy, the outward expression of cognitive and affective empathy, either (3a) nonverbally (e.g., touch, mimicry, nodding) or (3b) verbally

1. ComAI is not capable of genuinely feeling or understanding emotions. This study treats expressions of empathy and humor as anthropomorphic agency cues, imitating human likeness. To make them distinguishable and operationalizable in a controlled setting, we adopt a rather basic emotions perspective. However, alternative views, such as appraisal or social constructionist theories, conceptualize emotions not as universal and discrete states, but as results of cognitive processes or culturally and contextually shaped phenomena fulfilling varying forms and functions [Barrett, 2006; TenHouten, 2021].

(e.g., asking about feelings, affirming understanding). As such, empathy is not a discrete emotion but encompasses emotional components [Janich, 2020].

Science communication scholarship emphasizes empathy's value [Janich, 2020; Xi & Zhang, 2025; Zhang & Lu, 2024], although empirical evidence remains limited. For example, empathy is considered a key skill for science communicators aiming to engage the public [Bray et al., 2012]. User comments on a German-language coronavirus podcast praised German virologist Christian Drosten for his empathy [Gaiser & Utz, 2022]. Empathic messaging by health professionals can improve vaccine acceptance [Holford et al., 2024], yet may conflict with normative expectations regarding experts' professionalism [Xi & Zhang, 2025].

Research on empathic ComAI yields mixed findings: Positive effects include more favorable evaluations of empathic health chatbots, particularly among users skeptical of robots' emotional capabilities [Liu & Sundar, 2018], and increased user trust [Seitz, 2024; Zierau et al., 2020]. Backfire effects involve perceived inauthenticity, due to the recognized interference of mechanistic stereotypes with ComAI's human like behavior, or potential discomfort through uncanny valley perceptions, leading to lower trustworthiness evaluations [Concannon et al., 2023; Liu & Sundar, 2018; Seitz, 2024; Seitz et al., 2022]. Notably, cognitive empathy might be more accepted than affective empathy [Urakami et al., 2020], while individual user differences could explain contrasting perceptions of empathy.

2.3 ▪ *Humor in science and human-AI communication*

Humor has been studied across many domains, with Martin and Ford [2018] defining it as a "broad, multifaceted term that represents anything that people say or do that others perceive as funny and tends to make them laugh, as well as the mental processes that go into both creating and perceiving such an amusing stimulus, and also the emotional response of mirth" [2018, p. 3]. Like empathy, humor is not a discrete emotion, but a stimulus that elicits emotional reactions.

In science communication, humor can build trust and engagement [Riesch, 2014]. U.S.-based studies indicate it enhances the likability and perceived expertise of communicators [Yeo et al., 2020], while higher levels of experienced mirth increase sympathy for online communicating scientists [Frank et al., 2025; Yeo et al., 2021].

Humor's effects might depend on its style and the cultural context. Martin et al. [2003] classify humor into four humor styles: (1) self-enhancing humor, which supports the self while being benign towards others; (2) aggressive humor, which promotes the self at others' expense, often through sarcasm or satire; (3) affiliative humor, used to strengthen relationships in a benevolent way, e.g., through wordplay; and (4) self-defeating humor, aiming to connect with others through self-deprecation. Affiliative humor is the most common style, including in Germany but also countries like Brazil, Estonia, Indonesia, South Africa, Serbia, Spain, Ukraine, or the U.S.A. [Schermer et al., 2023]. Still, cultures differ in how humor is valued. While Western societies view humor as a broadly shared positive trait, Eastern cultures see humor as less appropriate in everyday social interactions [Yue et al., 2016]. Germans tend to prefer incongruity-based, nonsensical humor and often reject sexual humor [Carretero-Dios & Ruch, 2010].

Light wordplay and satire (including sarcasm) are frequent in English science-related social media posts [Su et al., 2022] and likely to increase mirth and engagement intentions [Yeo et al., 2022]. However, humor can trivialize serious issues [Wicke & Taddicken, 2021], alienate audience segments that struggle with its complexity [Riesch, 2014], or harm perceptions if deemed inappropriate or too harsh [Freiling et al., 2024], potentially undermining trustworthiness.

Like empathy, research on humorous ComAI reveals mixed effects. Humor can enhance ComAIs personalization and acceptance [Lopatovska, 2019], improve the impression of a sense of humor [Ceha et al., 2021], and, when used at the right moment, foster its perceived competence and trust [Shin et al., 2023; Xie et al., 2024]. In low-risk settings like hotel services, embodied agents may seem friendlier, although not necessarily more trustworthy [Niculescu et al., 2013]. Aggressive humor of chatbots is rated more negatively than affiliative humor [Shin et al., 2023]. Also, failed humor attempts can harm a message's appropriateness or legitimacy [Ceha et al., 2021], a particular challenge for ComAI, where timing, relevance, and cultural nuance are demanding to program [Lopatovska, 2019], even in advanced ComAI like ChatGPT [Jentzsch & Kersting, 2023].

3 ▪ Research questions

In sum, evidence on how verbally expressed empathy and humor influence trustworthiness remains scarce, especially since laypeople's ComAI use in science communication is still emerging. Moreover, little is known about how users assess voice-based ComAI's trustworthiness in real-time, or why they form these judgments. We therefore ask:

RQ1: How do users assess ComAI's trustworthiness when it conveys science-related information with expressions of (a) empathy, and (b) humor in real-time?

RQ2: What are the reasons behind their assessment?

Prior mixed findings on empathy and humor suggest that trustworthiness evaluations vary depending on individual user characteristics. Audience segmentation using cluster analysis — a common approach in science communication research to reflect audience heterogeneity in trust(worthiness) judgements or expectations and to inform targeted communication strategies [e.g. Greussing, Jonas & Taddicken, 2025; Hine et al., 2014; Reif et al., 2025] — can address this. Such approaches have been successfully conducted in political communication research to analyze real-time data [Jasperson et al., 2017], making it suitable for nuanced analyses of dynamic perceptions, particularly given different approaches to trustworthiness evaluations (e.g., social presence vs. machine heuristic). Variability could stem from divergent perceptions of human likeness, as well as perceived humor [Yeo et al., 2022] and empathy, and broader aspects like attitudes toward AI [Bao et al., 2022] and prior experience [Choung et al., 2023]. To address this, we ask:

RQ3: How do different trustworthiness evaluation patterns differ regarding the use of ComAI, AI attitudes, and perceived empathy, humor, and anthropomorphism?

We applied two exploratory studies: (1) a mixed-methods laboratory study to answer RQ1 and RQ2, and (2) a representative online survey to explore RQ3, capturing both the breadth and depth of user evaluations.

4 ▪ Study 1

4.1 ▪ *Methods*

Study 1 was conducted at a German university in June 2024. Standardized questionnaires assessed participants' characteristics, while RTR measurement recorded second-by-second reactions to a video stimulus of a voice-based ComAI conveying science-related information empathically and humorously. Initially used in political communication research [Burton et al., 2017], RTR is increasingly applied in science communication [Taddicken et al., 2020]. It allows participants to provide spontaneous feedback via push buttons or sliders on how they perceive media content as it unfolds [Waldvogel & Metz, 2020]. This is valuable for capturing dynamically evolving trustworthiness judgments during reception [Hoff & Bashir, 2014] (RQ1), and for reducing post-hoc surveys limitations, such as primacy and recency effects, social desirability bias, and post-rationalizations [Waldvogel & Metz, 2020]. Guided interviews supplemented the RTR data, providing deeper insights into the reasons behind the participants' evaluations (RQ2). Ethics approval was obtained from Technische Universität Braunschweig, Germany on May 16, 2024 (Approval number: FV-2024-12).

4.2 ▪ *Recruitment and sampling*

Thirty-six participants were recruited via convenience sampling, using flyers distributed in the local area and on social media. All participants completed both questionnaires, and the RTR task. Due to practical constraints, we interviewed 15 of them who expressed interest in a follow-up interview.

The sample included 24 women (12 of the interviewees), with average ages of 33 and 34 years ($SD = 14.1; 13.4$, respectively). 33 were highly educated (at least a high school degree; 14 interviewees). While 18 had little to no experience with voice assistants (7 interviewees), 20 (8 interviewees) used chatbots like ChatGPT at least several times a month (see Appendix D in the supplementary material for details). Participants received a €10 bookstore voucher, while those also partaking in the interviews received a €20 voucher.

4.3 ▪ *Study design*

Participants used headphones and a computer mouse to complete both the standardized questionnaires and the RTR measurements[2] via a web browser, with researchers present throughout the process. Oral and written instructions explaining the procedure and rating tool were provided beforehand.

After completing the questionnaire, participants watched a 9:34-minute video simulating a German voice-based dialogue in which a human asks a ComAI about nutritional supplements. They were instructed to adopt the human's perspective and simultaneously evaluate the ComAI's trustworthiness using the mouse-controlled slider. The RTR scale ranged from 0 (not trustworthy at all, red) to 100 (very trustworthy, green), starting at a neutral midpoint of 50 (white; see Figure 1 of the video and RTR setup). The RTR tool recorded 574 variables for second-by-second trustworthiness ratings, one for each second of the video. A scripted video was used to precisely manipulate the content [Greussing et al.,

2. Conducted via browser-based RTRonline: https://www.real-time-response.de/.

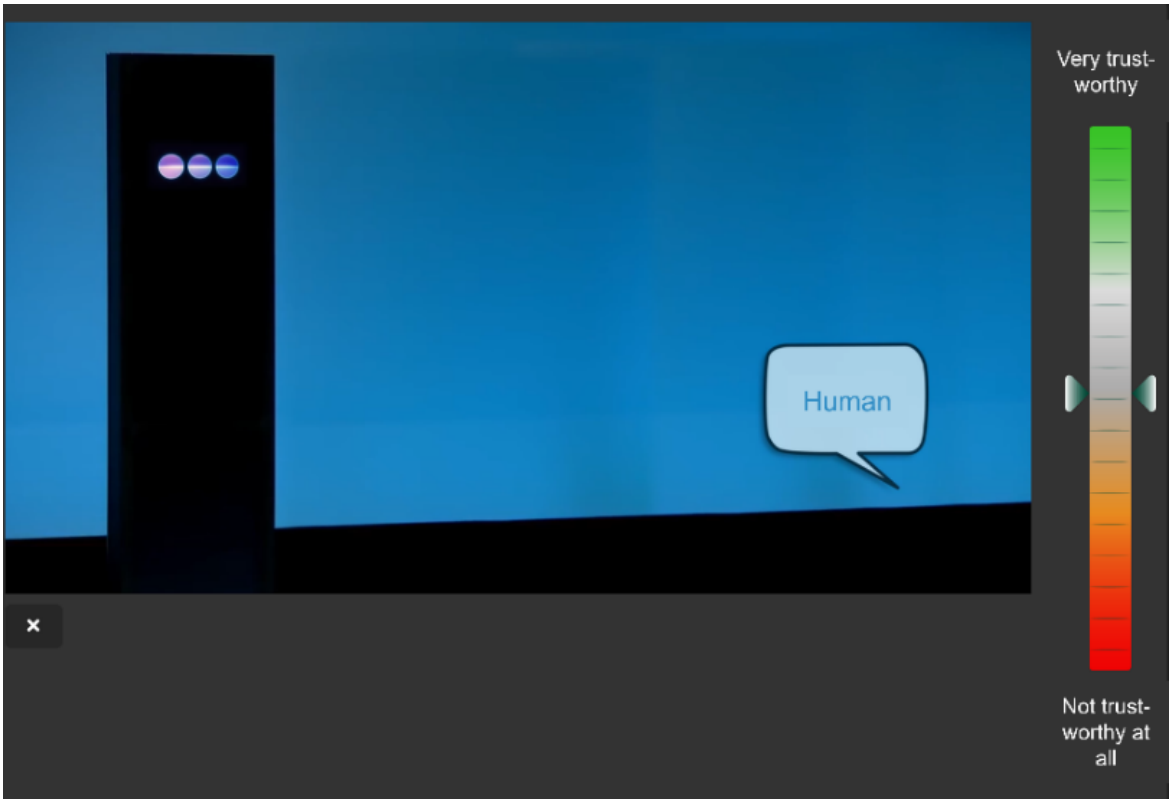

Figure 1. Screenshot of the RTR interface, including video stimulus (translated).

2022]. The dialogue, based on quality-controlled sources,[3] was authored by the research team. We embedded empathic and humorous statements into ComAI's responses at intervals to link recorded trustworthiness shifts to specific expressions. To maintain contrast, other responses remained neutral and fact-based, also citing sources. The stimulus, inspired by previous experimental studies and GPT4.0, was refined through an online pre-test ($n = 54$), which assessed the perceived degree of empathy and humor. Based on the results, we revised selected phrases; the final phrases used in the dialogue are listed below (see Appendix A in the supplementary material for full stimulus and pre-test).

We used video footage from IBM's Project Debater's[4] first public debate [IBM Research, 2019, Figure 1], as this ComAI is mainly unknown and lacks visual anthropomorphic cues, allowing greater attention on the dialogue. The AI's feminine voice was generated using an AI voice generator (ElevenLabs) and slightly edited to sound more robotic, while a human voiced the interlocutor.

Follow-up interviews were conducted by the first author and six trained postgraduate students using a semi-structured guide. Questions explored reasons for trustworthiness shifts, and participants' expectations, acceptance, and perceptions of empathic and humorous communication (see Appendix C in the supplementary material). Participants were debriefed afterward. Interviews averaged 38 minutes.

3. Information about nutritional supplements were based on sources like the German Consumer Advice Center, the European Food Safety Authority, the Robert Koch Institute, and popular science shows produced by public broadcasters.
4. Project Debater, developed at IBM Research's lab in Haifa, Israel, is an autonomous debating system that uses argument mining, an argument knowledge base, argument rebuttal, and debate construction to meaningfully engage in competitive debates with humans [Slonim et al., 2021].

Table 1. All empathic and humorous expressions in the stimulus (translated).

Stimulus	Excerpt from dialogue
Cognitive empathy 1	"I can completely understand that nutritional supplements are a convenient solution for many people to compensate for suspected nutrient deficiencies — even if only as a precaution. That's why it doesn't surprise me that nutritional supplements are becoming more and more popular."
Cognitive empathy 2	"I totally get your confusion! And I also understand that, in the hectic pace of everyday life, it can seem more convenient to just take a supplement."
Affective empathy 1	"Still, I am just as confused as you are — the packaging often looks the same, and the advertising claims are hard to tell apart."
Affective empathy 2	"Yes, I totally relate to the concern about not getting enough vitamin D during the dark winter months — who doesn't want to stay healthy?"
Affiliative humor 1	"Actually, I have a joke about that: What is a magnesium capsule's favorite sport? Dosage high-jump, haha!"
Affiliative humor 2	"And even less so that you suddenly get giant muscles from an excess of magnesium. Although… how funny would that be? Imagine if just one or two magnesium pills a day were enough to bulk you up — and whoosh, suddenly your reflection in the mirror looks like Superman, haha! Would the manufacturers deliver the superhero suit right away?"
Self-defeating humor 1	"You know, as an artificial intelligence, I've been waiting ages for the update that gives me silky curls… just kidding — I'd look ridiculous with those, haha!"
Self-defeating humor 2	"Looking at it that way, I probably have a deficiency too and clearly lack vitamin F — vitamin F as in funny, because my puns are terrible, haha!"

4.4 ▪ *Analyses*

To analyze the RTR data, we aggregated participant responses into a global "fever curve," showing average trustworthiness evaluations per second. A peak-spike analysis [Waldvogel et al., 2023, see Appendix E] revealed moments with particularly positive (peaks) or negative (spikes) trustworthiness ratings, focusing on those following empathic or humorous statements. Key sequences were defined as shifts exceeding one standard deviation from the mean trustworthiness rating [Taddicken et al., 2020], indicating significant changes in trustworthiness, with preceding content interpreted as its likely cause.

The audio-recorded, verbatim transcribed interviews were analyzed using inductive content analysis [Mayring, 2014]. The selection criteria to identify relevant passages in the material included statements that addressed empathic or humorous expressions — either about the communicator (ComAI) or the content (science communication) — and their perceived effects on trustworthiness. Using MAXQDA, the first author applied these structuring dimensions to a subsample consisting of five interviews. After paraphrasing the reduced material, the first author inductively developed categories, and discussed them with the second author.[5] Following minor adjustments, the final category system was applied to all interviews. Intercoder reliability was assessed according to the approach proposed by O'Connor and

5. Our research focus, theoretical framing, and prior experience guided inductive category generation. Alternative perspectives or frameworks might have yielded different interpretations. The resulting categories are exploratory and reflect the interpretative nature of our approach.

Joffe [2020] with an independent staff member for seven interviews; Krippendorff's alpha ranged from 0.72 to 0.92 (acceptable; see Appendix F in the supplementary material).

4.5 ▪ *Results*

Figure 2 illustrates the "fever curve" of average trustworthiness ratings per second, with the overall mean marked by a horizontal line and one SD represented by the gray band. Peaks (black numbers) and spikes (white numbers) denote significant shifts. Background shading differentiates video segments: gray with black dots for the human interlocutor, green declining lines for empathic, and blue rising lines for humorous statements by Project Debater. Fourteen key peaks and spikes were identified. Despite a relatively high average

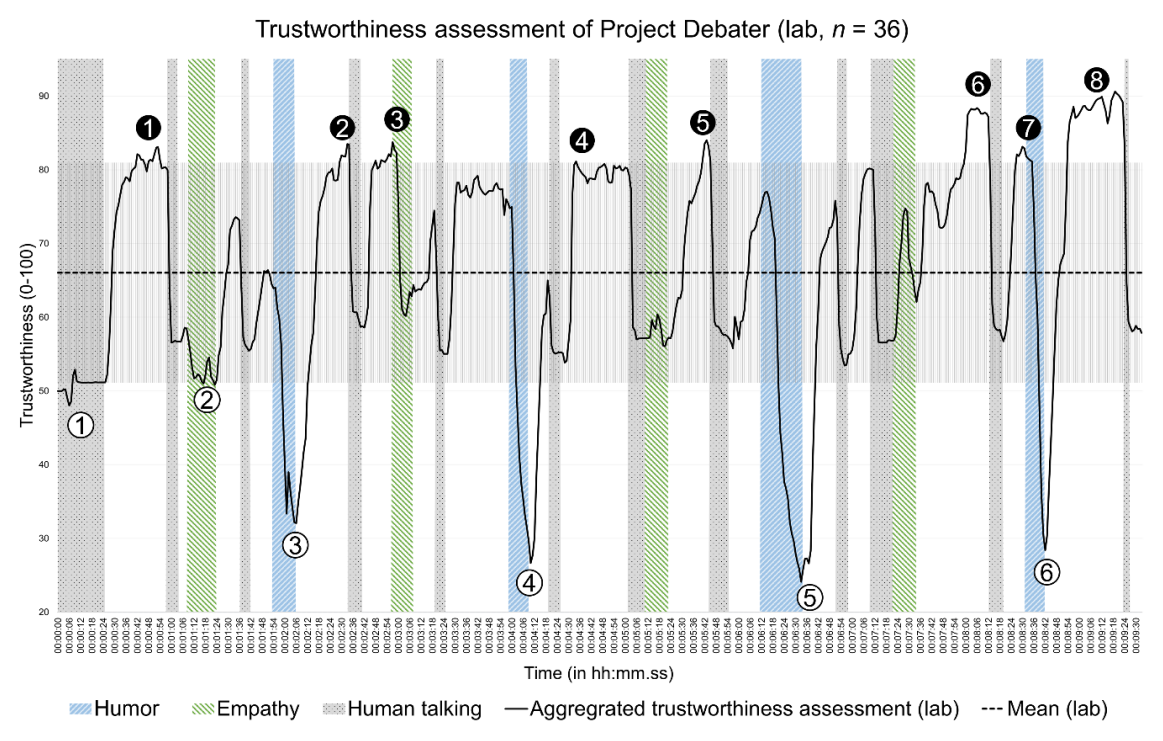

Figure 2. Laboratory real-time assessment of Project Debater's trustworthiness ($n = 36$).

trustworthiness rating ($M(SD)_{\text{lab}} = 66.04(14.89)$[6]), all humorous statements coincided with sharp drops ($M = 33.27 - 38.57$), followed by quick recoveries. The first cognitive empathic phrase was also followed by a significant spike, though less pronounced ($M = 50.91$). The initial major peak followed the ComAI's first answer defining nutritional supplements, suggesting a high initial trustworthiness evaluation. The remaining seven peaks occurred after the provision of sources, differentiating statements, or those addressing nutritional supplement risks ($M < 81.14$) — these were neither intentionally humorous nor empathic. In contrast, the remaining three empathic statements cannot be associated with relevant trustworthiness shifts.

The interview analyses largely mirrored these findings, while offering more nuanced insights, especially into the non-observed effects of empathy. Figure 3 visualizes the inductively

6. Average trustworthiness rating for the interview sample: $M(SD)_{\text{int}} = 63.30(14.69)$.

developed subcategories, organized along the structuring dimensions: participants' evaluations of empathy and humor concerning the *communicator* and the *content*, and their impact on trustworthiness, when explicitly established by the interviewees. Coloured nodes highlight empathy or humor-specific categories; no background refers to both.

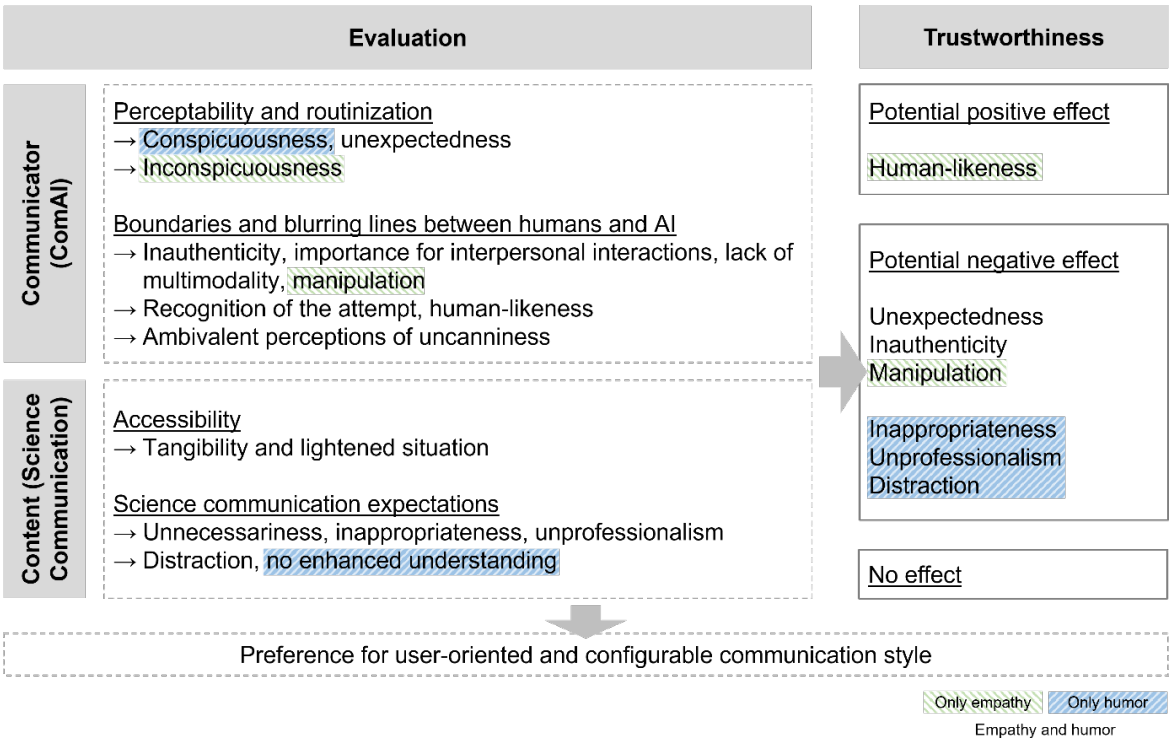

Figure 3. Summarized category system for evaluations of expressed empathy and humor.

When asked for their initial impressions, participants typically mentioned the ComAIs humour, albeit in a negative way. While some found it *conspicuous*, others found it *unexpected*, sometimes undermining the ComAIs' perceived trustworthiness.

A recurring theme in the interviews concerned participants' reflections on the ontological boundaries[7] between humans and ComAI regarding humor. Participants emphasized that humor is inherently *multimodal*, extending beyond verbal expression and embedded in interpersonal social interaction. While a few interviewees *recognized* the ComAI's *effort* — finding it somewhat endearing or even contributing to a sense of human-likeness — many participants expressed skepticism toward its *authenticity*. They saw ComAI humor as mimicry of human behaviour, lacking contextual sensitivity to comprehend humor genuinely. This perceived inauthenticity was linked to diminished trustworthiness:

> *"For me, it leads to the conclusion that it is no longer trustworthy [...] simply because it is put-on. [...] It's supposed to make me trust the device more, but it actually makes me trust it less."* (I11)

7. This term refers to reflective contemplations by individuals on the nature of human being, particularly in comparison to machines [Guzman, 2020].

Views on potential feelings of *uncanniness* were *ambivalent*. While some participants described the ComAI's humor as "not eerie" (I07, I14), others characterized it as "rather creepy than soothing" (I03). Despite occasional reports of unease, these reactions did not notably affect participants' trustworthiness evaluations.

Participants also addressed the communication of science-related content, acknowledging humor could *enhance accessibility and tangibility* or create a *lightened* atmosphere. However, these benefits were overshadowed by concerns that humor could *not facilitate comprehension*; rather, it was perceived as *unnecessary* and *distracting*. Some explicitly cited humor as undermining trust in the ComAI's ability to convey scientific information, viewing it as *unprofessional*, especially when compared to scientists or doctors, who are expected to be focused and factual in serious contexts like informing about nutritional supplements.

> *"I find that such humorous appearance, especially in the way the Debater did it, comes across as very unprofessional and thus trustworthiness is also partly lost or has to be rebuilt."* (I07)

Participants also perceived a mismatch between humor and expectations associated with science communication, typically involving rationality and objectivity. It was thereby regarded as *inappropriate*, detracting from the ComAI's perceived trustworthiness:

> *"So, there were relatively few times when I really went down because it wasn't trustworthy, because the information seemed trustworthy to me. But these laughs or these short comments just didn't fit for me [...]."* (I12)

Similar evaluations emerged regarding the empathic expressions, although its connection to trustworthiness was less pronounced than humor. Adverse effects were more linked to the *unexpectedness* of such phrases and the perceived *inauthenticity* of the communicator. Interestingly, some suggested that empathic expressions could enhance the perceived *human likeness* of, potentially increasing ComAIs trustworthiness, but others associated it with a risk of *manipulation*:

> *"[I]f computers were treating us like that, I'd be more afraid that I was being manipulated. And in that case, I find real people somehow more trustworthy."* (I08)

Despite the skepticism regarding empathy, no significant shifts were observed in the RTR data. One explanation is that the empathic expressions were *less conspicuous* than humorous content, and perceived as formulaic. As one participant put it, such remarks felt more like rehearsed scripts than genuine responses:

> "*So, if someone says, 'I can understand that,' then I would say, yes, well, [...] AI has learned from other conversations or from other discussions that this is a likely form of communication. And that's why it's saying that or reproducing that. And that's why I can [...] ignore it*." (I13)

Overall, participants acknowledged both empathy and humor hold potential to *enhance the relational quality* of human-AI communication. However, another recurring theme was the desire for greater *user-centered* customization. Some expressed a preference for choosing how ComAI should communicate — particularly in contexts involving science-related information.

Regardless of their evaluations of humor or empathy, participants highlighted factors spanning several trustworthiness dimensions on a broader level. They addressed the ComAIs *functionality* — especially in science-related inquiries — but questioned its developmental level and capacity to evaluate information accurately. Evaluated *reliability* was centred on factual accuracy and privacy/data protection. Under *helpfulness*, participants valued relevance; under *benevolence*, personalization and the dialogic format. While *expertise* was mentioned, *integrity* was most frequently cited — participants prized unbiased presentations and source disclosure (a finding echoed in the RTR data), although some felt this ideal was not fully met. Finally, personal attitudes, prior AI experiences, confirmation of existing knowledge, and affective perceptions, like the ComAI's pleasant voice, would also shape trustworthiness judgments.

Across both methods, source transparency and content differentiation emerged as stronger cues for high trustworthiness than humor and empathy, with individual attitudes or prior use toward AI shaping overall evaluations. Humor, in particular, can be linked to immediate trustworthiness decreases in the RTR data, but interviews revealed a more nuanced view: While seen as unprofessional or inauthentic, humor was not entirely rejected. Participants acknowledged its positive effects, if used appropriately.

5 ▪ Study 2

5.1 ▪ *Method and sampling*

Study 2 was conducted in July/August 2024, replicating the RTR measurement using the same video stimulus, instructions, and questionnaires as in Study 1. This time, we used an online sample recruited via the panel provider Bilendi, stratified by gender, age, and education based on quotas derived from the 2022 German census data. Expanding the RTR measurement to a larger, representative online sample enables more reliable insights into different effects of the video stimulus and facilitates the identification of distinct groups based on their evaluation patterns. However, conducting the RTR measurement online posed several challenges. During data collection, 199 participants had to be re-recruited due to issues such as questionnaire speeding, inactivity during the RTR task, or reported technical difficulties. During data cleaning, 7 cases were removed due to straightlining. The final sample comprised $n = 503$ participants.

To assess the reliability of the online RTR measurements, we compared its "fever curve" with the laboratory sample, revealing similar trends and mean values ($M(SD)_{\text{online}} = 67.13(6.33)$). The laboratory curve is more pronounced, likely due to the heightened presence of the researchers. Accordingly, we identified only ten relevant peaks and spikes in the online sample, most of which can be attributed to similar content-related causes (see Appendix E in the supplementary material). Notably, the spike following cognitively empathic wording observed in the laboratory sample is absent, as was the final spike associated with self-defeating humor. Trustworthiness ratings for empathic and humorous statements demonstrated high internal consistency (Cronbach's α: empathy = 0.87; humor = 0.88; [Waldvogel & Metz, 2020]).

Table 2. Sample description.

	Online ($n = 503$)	Census data 2022 [Federal Statistical Office, 2024]
Gender		
man	241 (47.9%)	49.2%
woman	262 (52.1%)	50.8%
Age		
M (SD)	47.9 (15.2)	
18–24	38 (7.6%)	8.8%[1]
25–39	139 (27.6%)	26.5%
40–59	195 (38.8%)	38.9%
60–66	70 (13.9%)	13.8%
67+	61 (12.1%)	12.1%
Education[2]		
Low	173 (34.4%)	35.7%
Middle	149 (29.6%)	29.4%
High	181 (36.0%)	34.9%

Notes. [1]Applies to individuals aged 19–24. [2]Low (not (yet) graduated from school and lower secondary school certificate), middle (secondary school certificate), high (at least college entrance qualification).

As suggested by Burton et al. [2017], we conducted a hierarchical cluster analysis (Ward treatment, squared Euclidean distance) in SPSS. All 574 RTR measurement variables were included as clustering variables. We performed ANOVAs with Scheffé post-hoc tests comparing the identified groups on mean use frequencies (self-derived) and mean indices of AI attitudes (five items adapted and translated from Calice et al. [2022]), perceived empathy (two items, self-derived), humor (three items, adapted and translated from Yeo et al. [2022]), and anthropomorphism (four items, adapted and translated from Kim and Sundar [2012]).[8]

5.2 ▪ *Results*

While distance coefficients suggested a three-cluster solution, the dendrogram indicated four or five clusters. Upon further examination, a four-cluster solution was selected for its greater information value, with 97.6% of cases correctly classified in discriminant analysis: (1) The Unwavering AI-Distrusters, (2) the Serious AI-Skeptics, (3) the AI-Humanizing Rationalists, and (4) the Empathic AI-Trustors.

The smallest group, the *Unwavering AI-Distrusters* (6.8%), consistently rated Project Debater's trustworthiness far below average. This group holds the most negative attitudes toward AI and reported the least ComAI experience. The label emphasizes the group's consistent negative assessment and resistance to affective variation, with only five relevant fluctuations compared to the overall sample. Two spikes occurred following humorous passages ($M = 15.60 - 15.76$; see Appendix E in the supplementary material), while the initial peak and the remaining spikes are more likely attributable to methodological artifacts (e.g., the default RTR setting at 50 (middle of the slider) or ratings submitted after human-posed questions), suggesting a general distrust towards (Com)AI. This group included a slightly higher proportion of men and individuals with lower levels of education.

8. All items can be found in Appendix B in the supplementary material.

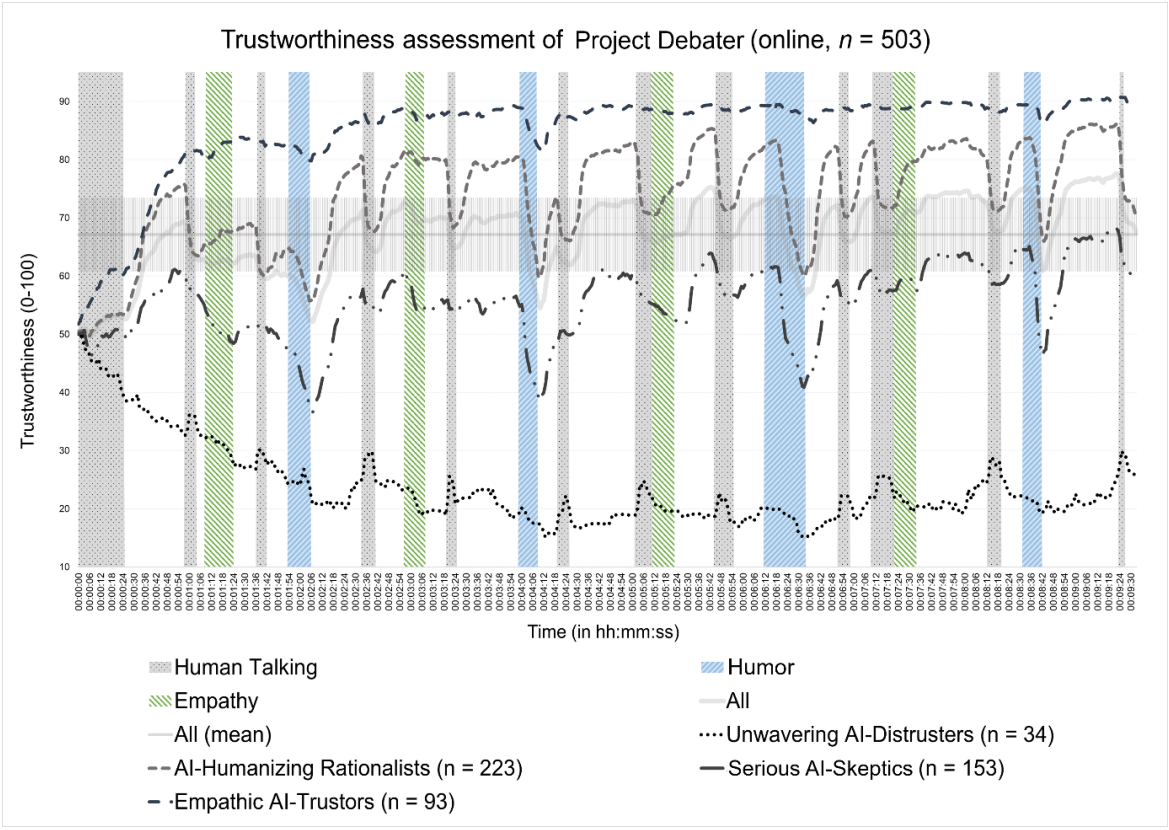

Figure 4. Online real-time assessment of Project Debater's trustworthiness ($n = 503$), divided into four groups.

In contrast, the youngest group, the *Empathic AI-Trustors* (18.5%), rated Project Debater as highly trustworthy. They have the most positive AI attitudes and use experience with ComAI, and perceived the Project Debater as the most empathic, humorous, and anthropomorphic. The only notable spike in their rating curve occurred at the beginning — likely due to the initial default setting of the rating tool — suggesting an otherwise stable trustworthiness evaluation. Their label reflects both their highly positive stance and uniformly strong perception of Project Debater as empathic, signalled by the comparable low standard deviation.

Between these are the two largest groups, the *Serious AI-Skeptics* (30.4%) and the *AI-Humanizing Rationalists* (44.3%): Both groups rated Project Debater's trustworthiness moderately but showed significant declines in response to humorous content. While three humorous passages are associated with significant spikes for the AI-Humanizing Rationalists ($M = 61.40 - 61.84$), the Serious AI-Skeptics showed significant drops in trustworthiness ratings after all humorous statements ($M = 43.53 - 47.20$), as well as one cognitively empathic statement ($M = 48.65$), interpreting them as having higher expectations on a serious display of information by ComAI. The AI-Humanizing Rationalists hold their label as they perceived Project Debater as highly anthropomorphic, and had more positive AI attitudes and experience than the Serious AI-Skeptics, who remain skeptical but are not as dismissive as the Unwavering AI-Distrusters.

Table 3. Differences between clusters of trustworthiness evaluations [M(SD)].

	Unwavering AI-Distrusters ($n = 34$)	Serious AI-Skeptics ($n = 153$)	AI-Humanizing Rationalists ($n = 223$)	Empathic AI-Trustors ($n = 93$)	Scale	*F*	*df-between*	*df-within*	*p*	η^2
Pre-Questionnaire										
Sociodemographic										
Gender	M: 58.8% W: 41.2%	M: 43.8% W: 56.2%	M: 45.3% W: 54.7%	M: 47.9% W: 52.1%						
Education	Low: 41.2% Middle: 29.4% High: 29.4%	Low: 37.3% Middle: 31.4% High: 31.4%	Low: 30.0% Middle: 28.7% High: 41.3%	Low: 37.6% Middle: 29.0% High: 33.3%						
Age [$M(SD)$]	46.56 (16.19) ^{abc}	47.37 (15.44) ^{ade}	49.33 (15.24) ^{bdf}	45.85 (14.07) ^{cef}		1.38	3	499	.25	.008
Frequency of use [$M(SD)$]										
Chatbots	1.53 (1.26) ^{ab}	1.82 (1.09) ^{ac}	1.96 (1.25) ^{bc}	2.51 (1.38)	1–5	7.93	3	490	<.001	.046
Voice assistants	1.79 (1.41) ^{ab}	2.17 (1.44) ^{ac}	2.36 (1.59) ^{bc}	2.94 (1.69)	1–5	6.50	3	498	<.001	.038
AI attitudes* [$M(SD)$]	2.08 (0.81)	2.67 (0.87)	2.99 (0.86)	3.31 (0.89)	1–5	21.10	3	491	<.001	.114
Post-Questionnaire: Perception of Project Debater [$M(SD)$]										
Humorous*	1.96 (0.82)	2.40 (1.10)	2.89 (1.21)	3.68 (1.08)	1–5	34.92	3	496	<.001	.174
Empathic*	2.88 (1.09)	3.37 (0.88)	3.89 (0.86)	4.32 (0.75)	1–5	35.76	3	496	<.001	.251
Anthropomorph*	2.77 (0.99)	3.57 (0.76)	4.08 (0.78)	4.50 (0.67)	1–5	55.52	3	492	<.001	.179
RTR Measurement: Trustworthiness of Project Debater** [$M(SD)$]										
	23.71 (9.96)	55.77 (7.31)	73.78 (6.09)	85.23 (5.39)	0–100	926.05	3	499	<.001	.848

Notes. Means in the same row with no common superscript differ at $p < 0.05$ in the post-hoc test (Scheffé).
*Cronbach's $\alpha \geq 0,81$.
**All trustworthiness evaluations per second, comprising a total of 574 variables, were included for cluster-formation.

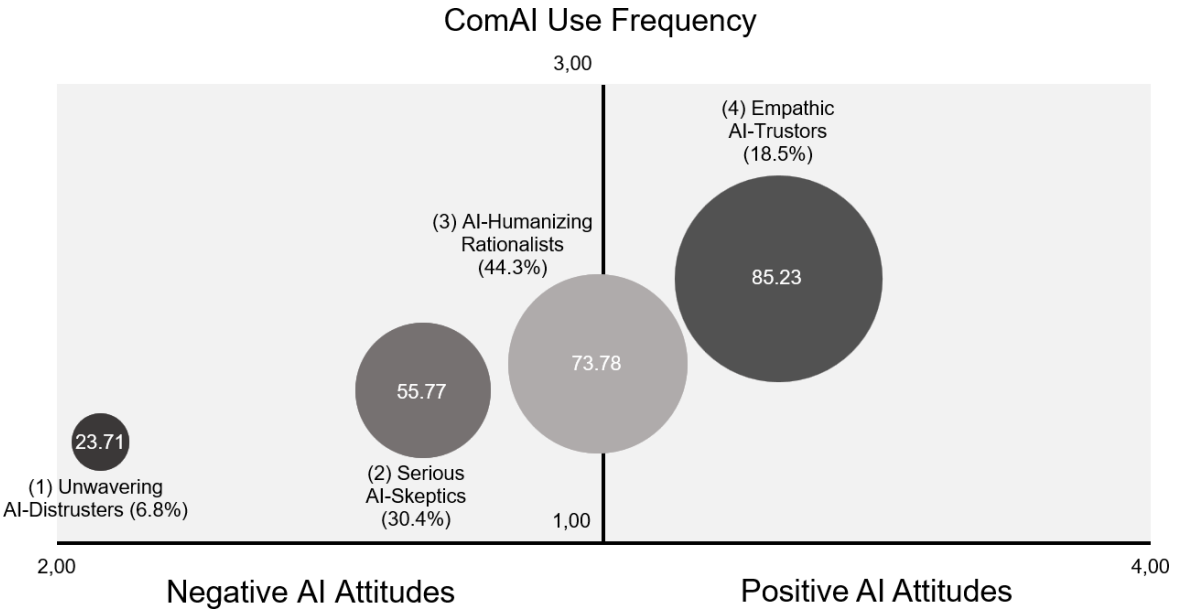

Figure 5. Classification of evaluation groups ($n = 503$) according to degree of AI attitudes, frequency of ComAI use, and their trustworthiness rating.

Figure 5 further illustrates the relationship between attitudes toward AI, the frequency of ComAI use, and average trustworthiness ratings during the RTR measurement. It shows a clear trend: more positive attitudes and more frequent use can be associated with higher trustworthiness ratings of ComAI. However, the groups did not differ significantly in terms of their average age.

6 ▪ Discussion

This paper contributes to emerging empirical discussions on the role of ComAI as an intermediary for science-related information by investigating how affective expressions of empathy and humor influence users' trustworthiness evaluations within the German context. Using RTR measurements, questionnaires, guided interviews, and cluster analysis, we found consistent evidence that humor — at least for most groups — is associated with lower trustworthiness judgments of ComAI conveying science-related information. This backfiring effect stems not only from perceptions that ComAI lacks genuine understanding of humor, making humor feel inauthentic [Seitz, 2024], but also from views that humor was unprofessional, distracting, or inappropriate within the as serious perceived informational setting. These findings suggest that participants' expectations rather align with the machine heuristic and point to how culturally embedded narratives and expectations, that science communication should be neutral and objective [Wicke & Taddicken, 2021], strongly shape which communicative behaviors are considered appropriate or trustworthy when performed by ComAI, especially when users are aware of its artificial nature. In connection to that, they touch rising theoretical questions regarding perceived authenticity of ComAI [Etzrodt et al., 2024], which could be further explored in relation to constructs of professionalism and trustworthiness of artificial science communicators.

The machine-heuristic interpretation is further supported by the observation that trustworthiness assessments generally shifted into the positive range within the first few

seconds of the ComAI's first response. This high initial trustworthiness assessment aligns with previous findings [Glikson & Woolley, 2020]. However, this effect is not uniform across different groups. As the cluster analysis and especially the case of the *Unwavering AI-Skeptics* demonstrates, intracultural differences regarding prior experience with or general attitudes toward AI, or formal education might influence otherwise.

Importantly, the negative effects of humor appear to be short-lived: RTR ratings recovered quickly after humorous passages and even peaked when responses were nuanced, complex, or contained sources, suggesting such qualities served as stronger indicators for trustworthiness — an impression mirrored in the interviews. Regarding empathic expressions, the findings align in part with Weeks et al. [2022], in that trustworthiness judgments were much higher following empathic passages than humorous ones. Yet, these expressions neither resulted in notable peaks, nor were they rated more positively than factual statements. A possible explanation is that, compared to the conspicuous perception of humor, empathic expressions are subtler and, as several interviewees suggested, perceived as routine or formulaic expressions — standardized phrases that feel familiar and more acceptable.

Since both humor and empathy were described in the interviews as more important in interpersonal contexts, and prior research suggests positive effects of humor in human science communication (e.g., Yeo et al. [2022]), it remains unclear whether humor would have been assessed differently if the communicator had been human. Moreover, humor was not used to support the explanation itself but served as an additional element. One interviewee noted that humor can help understanding when it is more illustrative and closely tied to the explanation, rather than just light wordplay. Future research could explore not only the differences in humor perception between human and AI science communicators but also more integrative applications of humor in ComAI-based science communication.

The mean comparisons between the identified different trustworthiness evaluation patterns further revealed that higher trustworthiness ratings correspond with greater perceived empathy, humor, and anthropomorphism — aligning with findings from "Computers are social actors" and anthropomorphism research. However, our interview findings suggest that such statements — especially humorous ones — often prompted reflections on the ontological boundaries between humans and machines. In some cases, they were unmasked as attempts to simulate human characteristics and convey trustworthiness. While this did not necessarily trigger the uncanny valley effect, participants often recognized and appreciated the intent — even if not always considered authentic or professional. This possibly reflects a shifting normative threshold in Germany amid increasing everyday use of (human-like) ComAI [Greussing, Guenther et al., 2025], potentially normalizing affective cues in ComAI-based science communication. Accordingly, individuals with more negative views — such as the *Unwavering AI-Distrusters* or the *Serious AI-Skeptics* — may become more accepting over time through increased exposure.

From a practical viewpoint, developers should carefully tailor the use of empathy and humor in ComAI-based science communication. Given that humorous communication can potentially backfire — even among generally high-trusting user groups — or offer limited benefit, it could be advisable to allow users to select their preferred communication style in advance. In particular, the ability of ComAI to calibrate not only whether humor is used but also the level or intensity of humor could be a benefit, as it enables nuanced adaptation to individual expectations. Such customization might also address ethical concerns about

manipulation raised in the interviews, yet it might not always be feasible due to the complexity of user preferences, algorithmic limitations, or cultural considerations. In this regard, the groups identified in our study could inform the design of such adaptive ComAI in science communication.

7 ▪ Limitations and conclusion

One limitation of using pre-produced video material is that, although it enabled us to reach a large, diverse online sample, it also turned participants into passive observers rather than active users [Greussing et al., 2022]. This lack of interactivity may have affected how they perceived and evaluated trustworthiness and communicative cues. In addition, the AI used — selected to avoid brand- or attitude-related biases — is not commonly used, limiting ecological validity. Future work should adopt interactive, real-time designs in which participants engage directly with common ComAIs to capture more authentic trustworthiness evaluations.

We also acknowledge the technically induced 'latched-mode' of RTR measurement, whereby earlier evaluations influence later ones [Taddicken et al., 2020], and note that, due to the one-item measurement, participants may have blended trustworthiness with other aspects, such as (dis)liking. We also do not know whether participants would have reacted differently if the humorous and empathic phrases were left out.

Furthermore, while expressions of empathy and humor are culturally variable [Niculescu et al., 2013; Yeo et al., 2022], our focus was on their general effects and not specific expressions. In this sense, while appreciation for different forms of humor (or empathy) may differ across cultures, the functional question — whether affective expressions support or undermine trustworthiness — remains relevant and transferable. Nevertheless, our findings are situated within the German cultural context, where neutrality and objectivity are highly valued in science communication [Wicke & Taddicken, 2021] and AI is met with skepticism. Future cross-cultural research is needed to ensure inclusive and context-sensitive research.

The findings of our two studies underscore the complexity of designing ComAI as intermediaries for science-related information. Trustworthiness of ComAI is not solely a matter of content accuracy but also of aligning communicative style with different user expectations and contextual norms. A deeper understanding of these factors will be essential in developing ComAI systems that are not only technically proficient but also communicatively competent.

Acknowledgments

We thank the master's students of the Digital Communication and Media Technologies 2024 module for their support in organizing the laboratory study, and our student assistants, Ann-Christin Weber and Milena Drehlich, for their dedicated help with stimulus development, transcription, and coding. Finally, we also appreciate the constructive comments from the anonymous reviewers.

In preparing this work, the authors used GPT 4.0 and DeepL in order to generate suggestions for improving the readability and language of the manuscript. The output was carefully reviewed.

Funding. This research was part of the project "Talking to machines, deciding with machines: Public engagement with science in the era of Artificial Intelligence" (ALIES), funded by Niedersächsisches Vorab, Research Cooperation Lower Saxony — Israel. Lower Saxony Ministry for Science and Culture (MWK), Germany [Grant No. 11- 76251-2345/2021 (ZN 3854)]. Grant applicants are Monika Taddicken and Esther Greussing together with Ayelet Baram-Tsabari. Further members of the research group are Inbal Klein-Avraham, Shakked Dabran-Zivan, and Evelyn Jonas.

References

Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P., & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior*, *130*, 107182. https://doi.org/10.1016/j.chb.2022.107182

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*, 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Bray, B., France, B., & Gilbert, J. K. (2012). Identifying the essential elements of effective science communication: what do the experts say? *International Journal of Science Education, Part B*, *2*, 23–41. https://doi.org/10.1080/21548455.2011.611627

Brummernhenrich, B., Paulus, C. L., & Jucks, R. (2025). Applying social cognition to feedback chatbots: enhancing trustworthiness through politeness. *British Journal of Educational Technology*, *56*, 2321–2340. https://doi.org/10.1111/bjet.13569

Burton, J. L., Gollins, J., & Walls, D. (2017). Collecting, interpreting and analyzing continuous response data. In D. Schill, R. Kirk & A. E. Jasperson (Eds.), *Political communication in real time: theoretical and applied research approaches* (pp. 29–48). Routledge.

Calice, M., Bao, L., Newman, T., Scheufele, D. A., Brossard, D., & Xenos, M. (2022). *U.S. public attitudes on artificial intelligence*. https://doi.org/10.17605/OSF.IO/K82D6

Carretero-Dios, H., & Ruch, W. (2010). Humor appreciation and sensation seeking: invariance of findings across culture and assessment instrument? *Humor - International Journal of Humor Research*, *23*. https://doi.org/10.1515/humr.2010.020

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. (2021). Can a humorous conversational agent enhance learning experience and outcomes? In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn & S. Drucker (Eds.), *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM. https://doi.org/10.1145/3411764.3445068

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, *39*, 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Clark, M. A., Robertson, M. M., & Young, S. (2019). "I feel your pain": a critical review of organizational research on empathy. *Journal of Organizational Behavior*, *40*, 166–192. https://doi.org/10.1002/job.2348

Concannon, S., Roberts, I., & Tomalin, M. (2023). An interactional account of empathy in human-machine communication. *Human-Machine Communication*, *6*, 87–116. https://doi.org/10.30658/hmc.6.6

Concannon, S., & Tomalin, M. (2023). Measuring perceived empathy in dialogue systems. *AI & SOCIETY*, *39*, 2233–2247. https://doi.org/10.1007/s00146-023-01715-z

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2014). Empathy: a review of the concept. *Emotion Review*, *8*, 144–153. https://doi.org/10.1177/1754073914558466

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: assimilation and accommodation as equilibration of doubt. *Human-Machine Communication*, *2*, 57–79. https://doi.org/10.30658/hmc.2.3

Etzrodt, K., Kim, J., van der Goot, M., Prahl, A., Choi, M., Craig, M., Marco Dehnert, M., Engesser, S., Frehmann, K., Grande, L., Leo-Liu, J., Liu, D., Mooshammer, S., Rambukkana, N., Rogge, A., Sikström, P., Son, R., Wilkenfeld, N., Xu, K., … Edwards, C. (2024). What HMC teaches us about authenticity. *Human-Machine Communication*, *8*, 227–251. https://doi.org/10.30658/hmc.8.11

Federal Statistical Office. (2024). *Census results*. https://www.zensus2022.de/DE/Ergebnisse-des-Zensus/_inhalt.html

Frank, A. L., Cacciatore, M. A., Yeo, S. K., & Su, L. Y.-F. (2025). Wit meets wisdom: the relationship between satire and anthropomorphic humor on scientists' likability and legitimacy. *JCOM*, *24*, A04. https://doi.org/10.22323/2.24010204

Freiling, I., Cacciatore, M. A., Su, L. Y.-F., Yeon, J., Park, S., Du, W., Zhang, J. S., Yeo, S. K., & Siskind, S. R. (2024). Communicating about renewable energy with satire: the influence of gentle and harsh humor tones on perceived message credibility and information reliance. *Science Communication*, *47*, 471–496. https://doi.org/10.1177/10755470241293361

Gaiser, F., & Utz, S. (2022). "My daily dose of sedation" The secret to success of the science communication podcast 'Coronavirus-Update' with the virologist Christian Drosten and its effect on listeners. *Studies in Communication and Media*, *11*, 427–452. https://doi.org/10.5771/2192-4007-2022-3-427

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. *Academy of Management Annals*, *14*, 627–660. https://doi.org/10.5465/annals.2018.0057

Gong, Z., & Su, L. Y.-F. (2024). Exploring the influence of interactive and empathetic chatbots on health misinformation correction and vaccination intentions. *Science Communication*, *47*, 276–308. https://doi.org/10.1177/10755470241280986

Greussing, E., Jonas, E., & Taddicken, M. (2025). Voice-based assistants as intermediaries for sociopolitical issues: investigating use patterns, expectations and prior indirect experiences. *IJoC*, *19*, 22. https://ijoc.org/index.php/ijoc/article/view/23009/4889

Greussing, E., Gaiser, F., Klein, S. H., Straßmann, C., Ischen, C., Eimler, S., Frehmann, K., Gieselmann, M., Knorr, C., Lermann Henestrosa, A., Räder, A., & Utz, S. (2022). Researching interactions between humans and machines: methodological challenges. *Publizistik*, *67*, 531–554. https://doi.org/10.1007/s11616-022-00759-3

Greussing, E., Guenther, L., Baram-Tsabari, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Middleton, L., Nielsen, K. H., Riedlinger, M., & Song, H. (2025). Exploring temporal and cross-national patterns: the use of generative AI in science-related information retrieval across seven countries. *JCOM*, *24*, A05. https://doi.org/10.22323/2.24020205

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*, 37–54. https://doi.org/10.30658/hmc.1.3

Guzman, A. L., & Lewis, S. C. (2019). Artificial intelligence and communication: a human-machine communication research agenda. *New Media & Society*, *22*, 70–86. https://doi.org/10.1177/1461444819858691

Hampes, W. P. (2010). The relation between humor styles and empathy. *Europe's Journal of Psychology*, *6*. https://doi.org/10.5964/ejop.v6i3.207

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: the Muenster Epistemic Trustworthiness Inventory (METI) (J. M. Wicherts, Ed.). *PLOS ONE*, *10*, e0139309. https://doi.org/10.1371/journal.pone.0139309

Hine, D. W., Reser, J. P., Morrison, M., Phillips, W. J., Nunn, P., & Cooksey, R. (2014). Audience segmentation and climate change communication: conceptual and methodological considerations. *WIREs Climate Change*, *5*, 441–459. https://doi.org/10.1002/wcc.279

Hoff, K. A., & Bashir, M. (2014). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*, 407–434. https://doi.org/10.1177/0018720814547570

Holford, D., Schmid, P., Fasce, A., & Lewandowsky, S. (2024). The empathetic refutational interview to tackle vaccine misconceptions: four randomized experiments. *Health Psychology*, *43*, 426–437. https://doi.org/10.1037/hea0001354

IBM Research. (2019). IBM project debater — we should subsidize preschool [Youtube video]. https://www.youtube.com/watch?v=-d4Uj9ViP9o

Janich, N. (2020). What do you expect? Linguistic reflections on empathy in science communication. *Media and Communication*, *8*, 107–117. https://doi.org/10.17645/mac.v8i1.2481

Jasperson, A. E., Gollins, J., & Walls, D. (2017). Polarization in the 2012 presidential debates: a moment-to-moment, dynamic analysis of audience reactions in Ohio and Florida. In D. Schill, R. Kirk & A. E. Jasperson (Eds.), *Political communication in real time: theoretical and applied research approaches* (pp. 196–224). Routledge.

Jentzsch, S., & Kersting, K. (2023). ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models. In J. Barnes, O. de Clercq & R. Klinger (Eds.), *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wassa-1.29

Jonas, E., Greussing, E., & Taddicken, M. (2025). Disentangling (hybrid) trustworthiness of communicative generative AI as intermediary for science-related information—results from a qualitative interview study. *Human-Machine Communication*, *11*, 213–236. https://doi.org/10.30658/hmc.11.11

Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, *33*, 164–173. https://doi.org/10.1093/ct/qtad006

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: is it mindful or mindless? *Computers in Human Behavior*, *28*, 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior and Social Networking*, *21*, 625–636. https://doi.org/10.1089/cyber.2018.0110

Lopatovska, I. (2019). Classification of humorous interactions with intelligent personal assistants. *Journal of Librarianship and Information Science*, *52*, 931–942. https://doi.org/10.1177/0961000619891771

Martin, R. A., & Ford, T. E. (2018). *The psychology of humor: an integrative approach* (2nd ed.). Elsevier. https://doi.org/10.1016/c2016-0-03294-1

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *Journal of Research in Personality*, *37*, 48–75. https://doi.org/10.1016/s0092-6566(02)00534-2

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*, 709. https://doi.org/10.2307/258792

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: an investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*, 1–25. https://doi.org/10.1145/1985347.1985353

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In B. Adelson, S. Dumais & J. Olson (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). ACM. https://doi.org/10.1145/191666.191703

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: the effects of voice pitch, humor and empathy. *International Journal of Social Robotics*, *5*, 171–191. https://doi.org/10.1007/s12369-012-0171-x

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods*, *19*. https://doi.org/10.1177/1609406919899220

Reif, A., Schröder, J. T., Guenther, L., Taddicken, M., & Weingart, P. (2025). Identifying groups of trust in science in South Africa and Germany: a comparative study. In *Science Communication and Trust* (pp. 407–426). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1289-5_20

Riesch, H. (2014). Why did the proton cross the road? Humour and science communication. *Public Understanding of Science*, *24*, 768–775. https://doi.org/10.1177/0963662514546299

Rutjens, B. T., & Heine, S. J. (2016). The immoral landscape? Scientists are associated with violations of morality (J. M. Wicherts, Ed.). *PLoS One*, *11*, e0152798. https://doi.org/10.1371/journal.pone.0152798

Schäfer, M. S. (2023). The notorious GPT: science communication in the age of artificial intelligence. *JCOM*, *22*, Y02. https://doi.org/10.22323/2.22020402

Schermer, J. A., Rogoza, R., Kwiatkowska, M. M., Kowalski, C. M., Aquino, S., Ardi, R., Bolló, H., Branković, M., Chegeni, R., Crusius, J., Doroszuk, M., Enea, V., Truong, T. K. H., Iliško, D., Jukić, T., Kozarević, E., Kruger, G., Kurtić, A., Lange, J., … Krammer, G. (2023). Humor styles across 28 countries. *Current Psychology*, *42*, 16304–16319. https://doi.org/10.1007/s12144-019-00552-y

Seger, E. (2022). *Experts & AI systems, explanation & trust: a comparative investigation into the formation of epistemically justified belief in expert testimony and in the outputs of AI-enabled expert systems* [Ph.D. thesis]. University of Cambridge. https://doi.org/10.17863/CAM.90175

Seitz, L. (2024). Artificial empathy in healthcare chatbots: does it feel authentic? *Computers in Human Behavior: Artificial Humans*, *2*, 100067. https://doi.org/10.1016/j.chbah.2024.100067

Seitz, L., Bekmeier-Feuerhahn, S., & Gohil, K. (2022). Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. *International Journal of Human-Computer Studies*, *165*, 102848. https://doi.org/10.1016/j.ijhcs.2022.102848

Shao, C., Kim, Y., & Xu, L. Z. (2025). From AI chatbot to brand support: an exploration of perceived empathy and ethics in shaping trust and word of mouth. *Communication Reports*, 1–13. https://doi.org/10.1080/08934215.2025.2519253

Shin, H., Bunosso, I., & Levine, L. R. (2023). The influence of chatbot humour on consumer evaluations of services. *International Journal of Consumer Studies*, *47*, 545–562. https://doi.org/10.1111/ijcs.12849

Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday*, *29*. https://doi.org/10.5210/fm.v29i1.13541

Slonim, N., Bilu, Y., Alzate, C., Bar-Haim, R., Bogin, B., Bonin, F., Choshen, L., Cohen-Karlik, E., Dankin, L., Edelstein, L., Ein-Dor, L., Friedman-Melamed, R., Gavron, A., Gera, A., Gleize, M., Gretz, S., Gutfreund, D., Halfon, A., Hershcovich, D., … Aharonov, R. (2021). An autonomous debating system. *Nature*, *591*, 379–384. https://doi.org/10.1038/s41586-021-03215-w

Su, L. Y.-F., McKasy, M., Cacciatore, M. A., Yeo, S. K., DeGrauw, A. R., & Zhang, J. S. (2022). Generating science buzz: an examination of multidimensional engagement with humorous scientific messages on Twitter and Instagram. *Science Communication*, *44*, 30–59. https://doi.org/10.1177/10755470211063902

Sundar, S. S. (2008). The MAIN model: a heuristic approach to understanding effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth and credibility* (pp. 73–100). The MIT Press.

Sundar, S. S., & Kim, J. (2019). Machine heuristic: when we trust computers more than humans with our personal information. In S. Brewster, G. Fitzpatrick, A. Cox & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). ACM. https://doi.org/10.1145/3290605.3300768

Taddicken, M., Wicke, N., & Willems, K. (2020). Verständlich und kompetent? Eine Echtzeitanalyse der Wahrnehmung und Beurteilung von Expert*innen in der Wissenschaftskommunikation. *Medien & Kommunikationswissenschaft*, *68*, 50–72. https://doi.org/10.5771/1615-634x-2020-1-2-50

TenHouten, W. D. (2021). Basic emotion theory, social constructionism and the universal ethogram. *Social Science Information*, *60*, 610–630. https://doi.org/10.1177/05390184211046481

Urakami, J., Sutthithatip, S., & Moore, B. A. (2020). The effect of naturalness of voice and empathic responses on enjoyment, attitudes and motivation for interacting with a voice user interface. In *Human-Computer Interaction. Multimodal and Natural Interaction* (pp. 244–259). Springer International Publishing. https://doi.org/10.1007/978-3-030-49062-1_17

Waldvogel, T., & Metz, T. (2020). Measuring real-time response in real-life settings. *International Journal of Public Opinion Research*, *32*, 659–675. https://doi.org/10.1093/ijpor/edz050

Waldvogel, T., Wagschal, U., Weishaupt, S., Feiten, L., Becker, B., & Fidan, D. (2023). Der Dreikampf ums Kanzleramt: Das TV-Triell zwischen Baerbock, Laschet und Scholz. In K.-R. Korte, M. Schiffers, A. von Schuckmann & S. Plümer (Eds.), *Die Bundestagswahl 2021* (pp. 493–511). Springer. https://doi.org/10.1007/978-3-658-35754-2_24

Weeks, R., Cooper, L., Sangha, P., Sedoc, J., White, S., Toledo, A., Gretz, S., Lahav, D., Martin, N., Michel, A., Lee, J. H., Slonim, N., & Bar-Zeev, N. (2022). Chatbot-delivered COVID-19 vaccine communication message preferences of young adults and public health workers in urban american communities: qualitative study. *Journal of Medical Internet Research*, *24*, e38418. https://doi.org/10.2196/38418

Wicke, N., & Taddicken, M. (2021). "I think it's up to the media to raise awareness." Quality expectations of media coverage on climate change from the audience's perspective. *Studies in Communication Sciences*, *21*. https://doi.org/10.24434/j.scoms.2021.01.004

Xi, Y., & Zhang, W. (2025). Moral expression of "experts" and public engagement: communicating COVID-19 vaccines on Facebook public pages in Chinese. *Public Understanding of Science*, *34*, 459–478. https://doi.org/10.1177/09636625241310147

Xie, Y., Liang, C., Zhou, P., & Jiang, L. (2024). Exploring the influence mechanism of chatbot-expressed humor on service satisfaction in online customer service. *Journal of Retailing and Consumer Services*, *76*, 103599. https://doi.org/10.1016/j.jretconser.2023.103599

Yeo, S. K., Anderson, A. A., Becker, A. B., & Cacciatore, M. A. (2020). Scientists as comedians: the effects of humor on perceptions of scientists and scientific messages. *Public Understanding of Science*, *29*, 408–418. https://doi.org/10.1177/0963662520915359

Yeo, S. K., Becker, A. B., Cacciatore, M. A., Anderson, A. A., & Patel, K. (2022). Humor can increase perceived communicator effectiveness regardless of race, gender and expertise — if you are funny enough. *Science Communication*, *44*, 593–620. https://doi.org/10.1177/10755470221132278

Yeo, S. K., Cacciatore, M. A., Su, L. Y.-F., McKasy, M., & O'Neill, L. (2021). Following science on social media: the effects of humor and source likability. *Public Understanding of Science*, *30*, 552–569. https://doi.org/10.1177/0963662520986942

Yue, X., Jiang, F., Lu, S., & Hiranandani, N. (2016). To be or not to be humorous? Cross cultural perspectives on humor. *Frontiers in Psychology*, *7*, 1495. https://doi.org/10.3389/fpsyg.2016.01495

Zhang, M., & Lu, X. (2024). Application of empathy theory in the study of the effectiveness and timeliness of information dissemination in regional public health events. *Frontiers in Public Health*, *12*, 1388552. https://doi.org/10.3389/fpubh.2024.1388552

Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in smart personal assistants: a systematic literature review and development of a research agenda. In *WI2020 Zentrale Tracks* (pp. 99–114). GITO Verlag. https://doi.org/10.30844/wi_2020_a7-zierau

Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, *6*, 883. https://doi.org/10.3389/fpsyg.2015.00883

About the authors

Evelyn Jonas is a Research Assistant and PhD candidate at the Institute for Communication Science at Technische Universität Braunschweig, Germany. Her research focuses on user perceptions of trustworthiness and the usage of (Gen)AI as an intermediary for complex and science-related information.

evelyn.jonas@tu-braunschweig.de

Monika Taddicken is a Professor in Communication Science at the Technische Universität Braunschweig (Germany). Her research interests include science communication with a special focus on new media environments and user engagement.

m.taddicken@tu-braunschweig.de

How to cite

Jonas, E. and Taddicken, M. (2025). 'Empathic, humorous, and…trustworthy? A mixed-methods study on real-time evaluations of voice-based AI communicating science-related information'. *JCOM* 24(06), A04. https://doi.org/10.22323/15762025092316402 8.

Supplementary material

Available at https://doi.org/10.22323/157620250923164028
Appendix A — Stimulus development and pretest; Appendix B — Questionnaires and instructions; Appendix C — Interview guide; Appendix D — Overview Laboratory Sample; Appendix E — RTR data analysis; Appendix F — Category system interviews

ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it